Peer Review

Review of: "Physical AI Agents: Integrating Cognitive Intelligence with Real-World Action"

Bahram Jalali[1]

1. University of California, Los Angeles, United States

Summary

High Level: This work extends the concept of Vertical AI Agents to “Physical AI Agents”, embodied AI with LLM cognition. The author proposes a standardized modular architecture for these Physical AI Agents with three core components (perception, cognition, and actuation) and introduces Physical Retrieval Augmented Generation (Ph-RAG) to connect physical intelligence with industry-specific LLMs. Case studies demonstrate applications across multiple sectors, including oil and gas pipeline monitoring and warehouse automation.

Detailed: The author first motivates the need for vertical AI agents. Since AGI is still undeveloped, the application of LLMs to specific complicated application verticals requires model fine-tuning and modification. They claim that adding in application-specific context, thereby creating “Vertical AI Agents”, can enhance performance in fields such as supply chain management, the oil and gas industry, healthcare, and manufacturing. Next, the author claims that some fields require physical interaction for full automation, requiring the emergence of “Physical AI Agents”. The author claims the main contributions of the work are to highlight the need for vertical AI, propose a standard for Physical AI Agents, introduce the concept of Physical Retrieval-Augmented Generation (Ph-RAG), and showcase industry application. The author then describes the core components of Vertical AI Agents, followed by the extension to Physical AI Agents. The author proposes a standard architecture for Physical AI agents composed of perception, cognition, and actuation modules. Each of these modules is discussed in detail, with the cognition block serving as the bridge between perception and physical actuation. Applications and embodiments across various industries are discussed. The author then

 doi.org/10.32388/UGYMVS

briefly discusses current industry efforts for developing platforms that may support Physical AI Agent development. Lastly, the author discusses two case studies that demonstrate the use of Physical AI Agents, along with the introduction of the Physical Retrieval-Augmented Generation (Ph-RAG) process. Ph-RAG combines the Physical AI Agent with an industry-specific LLM that may be queried for downstream reporting and informed decision-making. Conclusions and future directions are discussed.

Comments

Strengths

The work establishes an interesting direction forward for embodied AI with LLMs. Many applications and case studies are described, which effectively deliver both the motivation and description of the proscribed standard architecture for Physical AI Agents.

Weaknesses

- The paper would benefit from more explicit connections to established work in embodied AI, robotics, and physical reasoning. More comparisons with existing approaches would strengthen the positioning of this work. Why do LLMs provide value in this case over models like CNN? How does this compare to traditional control theory?
- Some technical details about the implementation of the three-block architecture remain underspecified, particularly regarding how the specialized LLMs are trained and fine-tuned for physical reasoning tasks and how they integrate with perception systems.
- Overall, the main crux of the problem, which the reviewer sees as the synthesis of cognitive reasoning, perception, and actuation, is not discussed in any detail. The standardization of architectural terminology is not sufficient in proscribing a technical path forward.
- The Physical Retrieval-Augmented Generation (Ph-RAG) process is underdescribed and only detailed in one case study. Again, the work is missing the technical detail behind this idea.

Suggested Revisions

- **Differentiate from Robotics**: More clearly articulate how Physical AI Agents differ from and improve upon traditional robotics frameworks and autonomous systems.
- **Provide Implementation Details**: Add more technical specifics about how the specialized LLMs are trained and optimized for resource-constrained physical platforms, and how the perception-to-action mapping is accomplished. Describe the details of Physical Retrieval-Augmented Generation (Ph-RAG).
- **Differentiate Physical AI Standard Model**: Add technical details for how your approach to architecture standardization differs from other approaches, or if it is completely novel, express why the breakdown is sufficient and optimal.
- **Address Limitations**: Include a dedicated section discussing the limitations of the current approach and potential challenges in implementation, including computational constraints, safety considerations, and deployment complexities.

Declarations

Potential competing interests: No potential competing interests to declare.